Reply Comments from

THE FUTURE OF PRIVACY FORUM

to

FEDERAL COMMUNICATIONS COMMISSION

Washington, D.C. 20554

WC Docket No. 16-106:

Protecting the Privacy of Customers of Broadband

and Other Telecommunications Services

Jules Polonetsky, CEO

John Verdi, Vice President of Policy

Stacey Gray, Legal & Policy Fellow

THE FUTURE OF PRIVACY FORUM[*†]
1400 I St. NW Ste. 450
Washington, DC 20005
www.fpf.org

July 6, 2016

[*] The Future of Privacy Forum (FPF) is a Washington, DC based think tank that seeks to advance responsible data practices.
[†] The views herein do not necessarily reflect those of our members or our Advisory Board.

Table of Contents

Executive Summary 1

1. Many Commenters Agree that the FCC's De-Identification Framework is Insufficient 2
2. The FCC Should Promulgate Rules that Respond to the Current Online Data Collection Ecosystem 4
3. Regulators Should Avoid Mandating Overly Rigid Standardized Privacy Notices 6
4. Distinguishing Sensitive from Non-Sensitive Data is Common in Data Privacy Law and Does Not Require Intrusive Methods 6

Conclusion 8

Infographic: *A Visual Guide to Practical De-Identification* 9

Executive Summary

The Federal Communications Commission (hereinafter FCC or Commission) has so far received more than 270,000 comments in response to its proposed rules regulating Internet Service Providers (ISPs),[1] including from individuals, academics, advocacy organizations, industry, and technical experts. In our own comments, filed May 27, 2016, we primarily called on the FCC to recognize the spectrum of data identifiability, and to establish a framework recognizing the utility and reduced privacy risks of non-aggregate de-identified data. This would allow uses of pseudonymous or not readily identifiable data for limited purposes subject to meaningful controls.[2] These comments can be found in the FCC's Electronic Comment Filing System and online at www.fpf.org.[3]

In response to the diverse opinions expressed by other organizations, we are filing this Reply Comment in order to note certain areas of broad agreement, and to respond to comments with which we disagree or offer a unique point of view. Specifically, in this Reply, we note that: (1) a substantial group of diverse commenters agree that the FCC's proposed de-identification framework is insufficient; (2) the FCC's rules should reflect an accurate, non-hypothetical understanding of current online data collection; (3) inflexible standardization requirements for privacy notices can stifle innovative transparency mechanisms; and (4) it is both practical and desirable to distinguish between degrees of data sensitivity and the contextual use of personal information.

In each of these areas, as well as in our earlier comments, we aim to ensure that the FCC has a clear understanding of online privacy, and has all of the facts necessary to craft practical and relevant rules that will elevate industry norms.

[1] Protecting the Privacy of Customers of Broadband and Other Telecommunications Services, 81 Fed. Reg. 23359 (proposed April 20, 2016) (to be codified at 47 CFR 64), at para. 7 (hereinafter, *Notice*). The proposed rules apply to Broadband Internet Access (BIAS) providers, which comprise a subset of ISPs, in these comments we refer to ISPs for the sake of convenience and clarity.

[2] Comments of the Future of Privacy Forum to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, *available at* https://ecfsapi.fcc.gov/file/60002089525.pdf (hereinafter, *FPF Comments*).

[3] *Id*.

1. Many Commenters Agree that the FCC's De-Identification Framework is Insufficient

In our comments of May 27, 2016, we urged the Commission to further develop or re-consider the binary approach inherent in its broad proposed definition of customer proprietary information ("PI"), all of which is treated as "personal" and subject to the strict rules, unless very highly aggregated, at which point no rules apply.[4] As we noted, this antiquated "personal or non-personal" framework stands in sharp contrast to leading government and industry guidelines with respect to de-identified data, including the FTC's guidance in this area.[5]

Instead, we urged the FCC to recognize that data exists on a spectrum of identifiability. We have described this in the attached infographic, *A Visual Guide to Practical De-identification*. The spectrum ranges from **personal data** (including explicitly personal, "potentially identifiable," and "not readily identifiable"), to **pseudonymous data** (including key-coded, pseudonymous, and protected pseudonymous), to **de-identified data** (including de-identified and protected de-identified), and finally to fully **anonymous data** (including anonymous and aggregated anonymous data, such as high-level statistical trends). As data moves along this spectrum, the application of rigorous de-identification methods—such as blurring, perturbation, and suppression[6]—as well as contractual controls and other safeguards, increasingly diminishes data utility but also reduces or eliminates privacy and security risks.

A substantial, diverse group of commenters agree: **the FCC's proposed approach to de-identified, non-aggregate data is insufficient.** For example, a number of organizations suggest that the FCC create exceptions for sharing de-identified data with third-party researchers, including to protect consumers from cyber threats.[7] Although some organizations have expressed skepticism over whether data can ever be 100% de-identified,[8] this misses the point that de-identification is not a one-size-fits-all approach. Instead, different techniques and different levels of identifiability may be appropriate under different circumstances, depending on other factors, such as the sensitivity of the underlying data, the purposes for which it is being shared, the retention periods, and the other administrative or technical safeguards in place.

Many organizations, even if they do not agree with FPF's views regarding the Commission's proposed rules, nonetheless recognize that de-identified data carry different privacy and security

[4] *FPF Comments* at 1–7, 29–30; *Notice* para. 154 et seq. (proposing to allow ISPs to use, disclose, and permit access to "aggregate customer PI" if the provider: (1) determines that the aggregated customer PI is not reasonably linkable to a specific individual or device; (2) publicly commits to maintain and use the aggregate data in a non-individually identifiable fashion and to not attempt to re-identify the data; (3) contractually prohibits any entity to which it discloses or permits access to the aggregate data from attempting to re-identify the data; and (4) exercises reasonable monitoring to ensure that those contracts are not violated).

[5] *FPF Comments* at 3.

[6] *Id.* at 6 (citing Simson L. Garfinkel, NISTIR 8053, *De-Identification of Personal Information* (Oct 2015), at 2, http://nvlpubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf).

[7] Comments of Nominum, Inc. to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 5, *available at* https://ecfsapi.fcc.gov/file/60002081097.pdf; Comments of William Lehr, Steve Bauer, & Erin Kenneally, to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 8–9, *available at* https://ecfsapi.fcc.gov/file/60002081123.pdf.

[8] *See, e.g.*, Comments of Privacy Rights Clearinghouse to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 5, *available at* https://ecfsapi.fcc.gov/file/60002081119.pdf.

impacts than data that has not been de-identified.[9] We agree, for example, with the Electronic Frontier Foundation, which noted that it might be appropriate to establish a different data retention standard for de-identified data due to its comparatively lessened risks.[10] We agree, too, with Consumers' Research, who point out that the Commission's strict binary approach eliminates regulatory incentives for companies to engage in de-identification efforts, resulting in the likelihood that fewer efforts at de-identification will be made.[11]

Others have pointed out that a number of useful de-identification techniques exist to address issues of privacy and security. The Electronic Privacy Information Center (EPIC), for example, acknowledges that an array of techniques for de-identification exist, noting that "because not all de-identification techniques adequately anonymize data, it is important that the process employed is robust, scalable, transparent, and shown to provably prevent the identification of consumer information."[12]

We agree that de-identification techniques vary in their effectiveness, and that more powerful techniques will reduce the risk of re-identification, while correspondingly reducing the utility of the data. For this reason, we disagree with the approach, suggested by one organization, of treating individualized de-identified data in the same manner as aggregated data, which the proposed rules would leave unregulated.[13] A wholesale exception for all stages of de-identified data would be similarly binary, and would fail to appropriately protect consumers from the risk, however minimal, of re-identification. For example, pseudonymous data may be considered de-identified, but making such data public or storing it without controls would be a privacy risk; that same data, subject to extensive controls and use limitations, may be stored and used with minimal risk. We simply propose that the FCC recognize that data can take many forms between explicitly personal and fully aggregated, and that some of these forms of data should be permitted to be used and shared subject to robust controls and consumer choices that are appropriate to its lesser privacy risks.

In our comments, we took particular note of the Federal Trade Commission's (FTC) extensive guidance regarding de-identification. According to the FTC, data are not "reasonably linkable" to

---

[9] *See* Comments of the Electronic Frontier Foundation to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 7, *available at* https://ecfsapi.fcc.gov/file/60002081036.pdf (hereinafter, *EFF Comments*); Comments of Consumers' Research to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 22–24, *available at* https://ecfsapi.fcc.gov/file/60002077807.pdf (hereinafter, *Consumers' Research Comments*); Comments of Access Now to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 11, *available at* https://ecfsapi.fcc.gov/file/60002078011.pdf ("BIAS providers should take all possible steps to ensure confidentiality of users. This includes anonymising information. While this technique is not perfect, it limits the retention period of this information to what is strictly necessary for a defined purposes and put data security measures in place to protect data integrity and prevent breach."); Comments of Verizon to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 44–45, *available at* https://ecfsapi.fcc.gov/file/60002078934.pdf.

[10] *EFF Comments* at 7 (noting, however, that such differential treatment could be complicated because "the risks of re-identification are difficult to assess and are known to increase over time as more data about individuals becomes available or as new re-identification techniques are developed.").

[11] *Consumers' Research Comments* at 22–24.

[12] Comments of the Electronic Privacy Information Center (EPIC) to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 21–22, *available at* https://ecfsapi.fcc.gov/file/60002079241.pdf (hereinafter, *EPIC Comments*). The technique that EPIC points out—differential privacy—is one of many techniques for de-identification, and we reiterate that many intermediate stages in this spectrum may be appropriate for certain types of data from ISPs in limited situations.

[13] *See Consumers' Research Comments* at 24; *Notice* at para. 154 et seq.

individual identity to the extent that a company: (1) takes reasonable measures to ensure that the data are de-identified; (2) publicly commits not to try to re-identify the data; and (3) contractually prohibits downstream recipients from trying to re-identify the data.[14] Industry self-regulatory guidelines use similar approaches.[15]

In their comments to the Commission, FTC staff note that:

> "[T]he [FCC's] proposal to include any data that is 'linkable' could unnecessarily limit the use of data that does not pose a risk to consumers. While almost any piece of data *could* be linked to a consumer, it is appropriate to consider whether such a link is practical or likely in light of current technology. FTC staff thus recommends that the definition of PII only include information that is 'reasonably' linkable to an individual."[16]

We echo and support the FTC's comments, and urge the FCC to recognize that non-aggregate data can be de-identified in a manner that makes it not reasonably linkable to a specific individual. This approach is consistent with leading government and industry guidelines with respect to de-identified data, including key work by the Federal Trade Commission, and is illustrated by FPF's *A Visual Guide to Practical De-Identification*.

2. The FCC Should Promulgate Rules that Respond to the Current Online Data Collection Ecosystem

We strongly believe that the FCC, in crafting rules to protect online consumer privacy, should have an accurate, **fact-based** understanding of the current online data ecosystem. For this reason, we dedicated a significant portion of our May 27th comments to explaining the current online ecosystem to contextualize the efficacy of the Commission's proposed approach to regulation of broadband privacy.

To that same end, we note that there has been an ongoing debate over the role of encryption and how it affects the visibility of personal information that ISPs are able to access about their customers.[17] Although an increasing percentage and volume of internet traffic is encrypted, organizations are divided on their understandings of how this affects (or should affect) regulation

---

[14] FEDERAL TRADE COMMISSION, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (2012), at 21, https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[15] *See* DIGITAL ADVERTISING ALLIANCE, SELF-REGULATORY PRINCIPLES FOR MULTI-SITE DATA (Nov 2011), at 8, *available at* http://www.aboutads.info/resource/download/Multi-Site-Data-Principles.pdf; NETWORK ADVERTISING INITIATIVE, 2015 UPDATE TO THE NAI CODE OF CONDUCT (2015), at 5, *available at* https://www.networkadvertising.org/sites/default/files/NAI_Code15encr.pdf.

[16] Comments of the Staff of the Bureau of Consumer Protection of the Federal Trade Commission (FTC) to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 9, *available at* https://ecfsapi.fcc.gov/file/60002078443.pdf (emphasis in original).

[17] *See, e.g.*, PETER SWIRE ET AL, INSTITUTE FOR INFORMATION SECURITY & PRIVACY, GEORGIA INSTITUTE OF TECHNOLOGY, ONLINE PRIVACY AND ISPS: ISP ACCESS TO CONSUMER DATA IS LIMITED AND OFTEN LESS THAN ACCESS BY OTHERS (May 2016), *available at* http://peterswire.net/wp-content/uploads/Online-Privacy-and-ISPs-1.pdf; UPTURN, WHAT ISPS CAN SEE (March 2016), *available at* https://www.teamupturn.com/reports/2016/what-isps-can-see.

of ISPs.[18] However, most commenters do not seriously debate the fact that encryption itself is an important privacy-enhancing tool.

Several commenters have remarked that a large amount of web surfing data is available even from sites that are encrypted.[19] With respect to claims that ISPs will use certain metadata, e.g. port numbers, domain names, and timing of web traffic, to support advertising, we are not aware of any ISP business model which relies on this information. In fact, the types of data that are available and being used for ad targeting today are quite visible and widely available, as we described in our comments.[20]

In our comments, we detailed the current role ISPs play in the advertising market today,[21] demonstrating how ISPs can support cross-device advertising by using state management capabilities to enable appending of third party data for use in ad targeting both in their own businesses and in partnership with third party ad networks. Specific examples of third party technology companies who currently work with ISPs around the world include ZeoTap, Bering Media and others.[22] These companies use de-identified data to enable privacy protective advertising that poses minimal risk to consumers and is typically subject to opt out controls. We suggest that this type of "privacy by design" business model, outlined extensively by former Ontario Privacy Commissioner Ann Cavoukian,[23] ought to be feasible under FCC rules.

One organization commented that the rise of predictive analytics in online marketing has not been well considered in this debate.[24] Predictive analytics and marketing are by no means new concepts in data use. The use of data to understand consumer preferences is a prime driver in many types of marketing and advertising. Ad targeting practices have leveraged a range of data sources to better understand consumer behavior since the earliest days of online advertising—similar to a range of parallel industries. Collecting data to better understand consumers and make marketing decisions is the very essence of the economics underpinning the Internet.

---

[18] *Compare, e.g.*, *EPIC Comments* at 23 (suggesting that FCC should mandate end-to-end encryption) *with* Comments of INCOMPAS to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 14, *available at* https://ecfsapi.fcc.gov/file/60002080517.pdf (suggesting that FCC should not mandate encryption).

[19] *See, e.g.*, Comments of Public Knowledge, the Benton Foundation, Consumer Action, Consumer Federation of America, and National Consumers League to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 6–11, *available at* https://ecfsapi.fcc.gov/file/60002080037.pdf (hereinafter *Public Knowledge Comments*); Comments of the Center for Democracy & Technology to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 16–17, *available at* https://ecfsapi.fcc.gov/file/60002079430.pdf; Comments of Consumer Watchdog to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 4, *available at* https://ecfsapi.fcc.gov/file/60002081032.pdf.

[20] *See FPF Comments* at 12–16 (describing the BlueKai database and the democratization of data).

[21] *FPF Comments* at 8–25 (Part II, "Explanation of the Market").

[22] *See* ZEOTAP, https://www.zeotap.com/ (last visited July 6, 2016); BERING MEDIA, http://www.beringmedia.com/ (last visited July 6, 2016).

[23] ANN CAVOUKIAN, REDESIGNING IP GEOLOCATION: PRIVACY BY DESIGN AND ONLINE TARGETED ADVERTISING (Oct 2010), *available at* https://www.ipc.on.ca/images/Resources/pbd-ip-geo.pdf.

[24] *Public Knowledge Comments* at 6–11.

3. Regulators Should Avoid Mandating Overly Rigid Standardized Privacy Notices

The Commission sought comment on different methods for "simplifying and standardizing privacy notices,"[25] including whether the FCC should require ISPs to create a consumer-facing "privacy dashboard."[26] As we discussed in our comments, we support robust and meaningful opt out mechanisms for consumers to exercise control over non-sensitive data, or data that is pseudonymous or not readily identifiable and therefore poses lessened privacy and security risks. We would urge the FCC to encourage industry and multi-stakeholder efforts to develop effective opt-out mechanisms, and to require ethics oversight for issues of fairness and anti-discrimination. We also support the comments of the International Association of Privacy Professionals (IAPP), which call for not just security training, but also privacy training, a key measure to ensure company employees are educated and informed about privacy requirements.[27] Taken as a whole, these measures will ensure meaningful controls and elevate current industry norms.

However, we would urge caution before the Commission implements a system of mandated and standardized privacy notices. While privacy law in the United States tends to center on disclosure by means of mandated notices, what we have seen is that current online platforms are constantly evolving in their privacy notices and choice mechanisms. In addition, some uses of data at issue may already be subject to widely used industry standard icons and notices.[28] As some privacy scholars have noted, mandated disclosures can be counter-productive if they do not provide relevant information about the practical consequences of data collection, or the practical benefits that may be lost if the data is not collected.[29] As a result, a system of mandated disclosures that is overly rigid or inflexible may be premature. Instead, this may be an area where the multi-stakeholder process would be more efficient to promote the development of more consistent and consumer-friendly privacy notices and choice mechanisms.

4. Distinguishing Sensitive from Non-Sensitive Data is Common in Data Privacy Law and Does Not Require Intrusive Methods

In our comments, we recommended that the FCC establish a multi-stakeholder process to develop privacy rules for sensitive ISP consumer data and out of context uses of such data. In our view, sensitive data would include the content of detailed browsing histories, as well as the more traditionally sensitive forms of data recognized by the industry today, such as health information

---

[25] *Notice* para. 58.
[26] *Notice* para. 95, 144–45.
[27] *See* Reply Comments of the International Association of Privacy Professionals (IAPP) to Federal Communications Commission (June 27, 2016), WC Docket No. 16-106, *available at* https://www.fcc.gov/ecfs/filing/1062787134854/document/10627871348546ec3.
[28] *See FPF Comments* at 27–28 (discussing the AdChoices icon and current self-regulatory requirements under the Network Advertising Initiative, and Digital Advertising Alliance).
[29] *See, e.g.*, Jane R. Bambauer et al, *A Bad Education*, U. ILL. L. REV. (forthcoming 2016), at 21–27, *available at* http://ssrn.com/abstract=2795808 (describing a theory of disclosure as a good or bad education); Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, U. CHI. L. & ECON., OLIN WORKING PAPER NO. 516, U. MICH. L. & ECON., EMPIRICAL LEGAL STUDIES CENTER PAPER NO. 10-008 (Mar. 2010), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1567284.

and precise geo-location. Other organizations have commented in a similar vein, arguing for the FCC to mirror the FTC's approach of distinguishing between sensitive and non-sensitive data.[30]

In contrast, some commenters have commented that sensitivity of data is difficult or impossible to assess, and as a result, all data should be treated as sensitive.[31] Specifically, one organization even commented that identification of sensitive data is impossible absent "intrusive methods."[32] This approach is at odds with a long history of data protection in the United States and worldwide. Under the recently approved EU General Data Protection Regulation (GDPR), categories of sensitive information are clearly enumerated.[33] Canada's PIPEDA similarly allows that the form of consent may vary depending on the sensitivity of the information.[34]

The FTC and the White House have similarly drawn distinctions between sensitive and non-sensitive data. In the 2012 Privacy Report, the FTC not only agreed that affirmative express consent is appropriate when a company uses sensitive data for any marketing,[35] but enumerated specific categories of information that ought to be considered sensitive.[36] The White House, in its most recent discussion draft of the Consumer Privacy Bill of Rights Act, proposes to require covered entities to provide individuals with "reasonable means to control the processing of personal data about them **in proportion to the privacy risk** to the individual," with privacy risk defined as the potential for the data to cause emotional distress, or physical, financial, professional

---

[30] *See, e.g.*, Comments of the Independent Telephone & Telecommunications Alliance (ITTA) to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 11–17, *available at* https://ecfsapi.fcc.gov/file/60002077236.pdf; Comments of CTIA to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 94–97, *available at* https://ecfsapi.fcc.gov/file/60002064853.pdf.

[31] *See Public Knowledge Comments* at 24–26 (alleging that determining sensitivity of data would require manual inspection of each packet); Reply Comments of Paul Ohm to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 2–6, *available at* https://ecfsapi.fcc.gov/file/10622254783425/OhmReplyComments.pdf.

[32] Comments of the National Consumers League (NCL) to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 2, *available at* https://ecfsapi.fcc.gov/file/60002078689.pdf.

[33] Regulation (EU) 2016/679, General Data Protection Regulation (GDPR), Art. 9 ("Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.").

[34] Personal Information Protection and Electronic Documents Act (PIPEDA), S.C. 2000, c. 5 (last amended 2015-06-23), Principle 4.3.5 ("In determining the form of consent to use, organizations shall take into account the sensitivity of the information. Although some information (for example, medical records and income records) is almost always considered to be sensitive, any information can be sensitive, depending on the context.").

[35] FEDERAL TRADE COMMISSION, REPORT: PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE (March 2012), at 47–48, *available at* https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf ("The Commission agrees with the commenters who stated that affirmative express consent is appropriate when a company uses sensitive data for any marketing, whether first- or third-party. Although, as a general rule, most first-party marketing presents fewer privacy concerns, the calculus changes when the data is sensitive. Indeed, when health or children's information is involved, for example, the likelihood that data misuse could lead to embarrassment, discrimination, or other harms is increased. This risk exists regardless of whether the entity collecting and using the data is a first party or a third party that is unknown to the consumer. In light of the heightened privacy risks associated with sensitive data, first parties should provide a consumer choice mechanism at the time of data collection. . . .").

[36] *Id*. at 59 ("Given the general consensus regarding information about children, financial and health information, Social Security numbers, and precise geolocation data, the Commission agrees that these categories of information are sensitive."). The framework of the FTC's Report was also determined to not apply to companies that collect only non-sensitive data from fewer than 5,000 consumers a year, provided they do not share the data with third parties. *Id*. at vii.

or other harm to an individual."[37] This and many other calculations of sensitivity occur throughout industry and governmental data protection regulations, belying the idea that sensitive data cannot be measured or regulated.

One commenter has pointed out that current industry self-regulatory guidance can vary in the types of data that are considered to be sensitive and thereby require opt in consent.[38] It is for that very reason that guidelines here, developed by a multi-stakeholder process, would be invaluable. Thus, we have called for the FCC to establish a multi-stakeholder process, led by the National Telecommunications & Information Association (NTIA). As discussed in President Obama's Consumer Privacy Blueprint, open, transparent multi-stakeholder forums can enable stakeholders who share an interest in specific markets or business contexts to work toward consensus on appropriate, legally enforceable codes of conduct.[39] In combination with limits on the uses of appended data, strict retention periods, and ethics oversight, we believe this would be the best way forward to regulate ISPs' uses of sensitive customer data.

With regard to the claim that identifying sensitive data will require intrusive methods, the FCC should take note of the range of methods used by ISPs that today or in the past have had reason to review web traffic. The claim that distinguishing sensitivity would require intrusive methods misses the point that the data at hand in many cases is data that has already been collected. In addition, ISPs can employ methods ranging from models that scan and do not log data other than whitelisted information, methods that scan and immediately delete (or not log at all) data that is identified as sensitive, or methods that log data but immediately categorize it as a high level rather than keep the underlying data. These established methods do not require companies to "manually inspect" the content of packets in order to avoid using sensitive data for targeted advertising.

Conclusion

For the above reasons, we reiterate our previous comments of May 27, 2016, and urge the FCC to adopt a relevant, fact-based approach to possible uses of de-identified data as well to issues of online data collection, predictive analytics, online marketing, and sensitive data. Specifically, we urge the Commission to (1) issue a rule that recognizes that de-identification is not binary, but that data exists on a spectrum of identifiability; (2) recognize that non-aggregate data can be appropriately de-identified; (3) establish a framework that allows ISPs to use data that are pseudonymous or not readily identifiable for limited purposes; and (4) establish a multi-stakeholder process to determine the best way to approach uses of data that are sensitive or out of context, taking into account degrees of identifiability.

---

[37] *See* WHITE HOUSE, ADMINISTRATION DISCUSSION DRAFT: CONSUMER PRIVACY BILL OF RIGHTS ACT OF 2015, *available at* https://www.whitehouse.gov/sites/default/files/omb/legislative/letters/cpbr-act-of-2015-discussion-draft.pdf.

[38] *See* Reply Comments of Paul Ohm to Federal Communications Commission (May 27, 2016), WC Docket No. 16-106, at 11–12, *available at* https://ecfsapi.fcc.gov/file/10622254783425/OhmReplyComments.pdf.

[39] THE WHITE HOUSE, CONSUMER DATA PRIVACY IN A NETWORKED WORLD: A FRAMEWORK FOR PROTECTING PRIVACY AND PROMOTING INNOVATION IN THE GLOBAL DIGITAL ECONOMY (February 23, 2012), at 2, http://www.whitehouse.gov/sites/default/files/privacy-final.pdf.

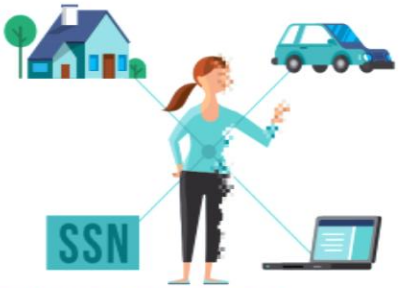
Infographic: *A Visual Guide to Practical De-Identification*

A VISUAL GUIDE TO PRACTICAL DATA DE-IDENTIFICATION

Produced by FUTURE OF PRIVACY FORUM FPF.ORG

In collaboration with EY

What do scientists, regulators and lawyers mean when they talk about de-identification? How does anonymous data differ from pseudonymous or de-identified information? Data identifiability is not binary. Data lies on a spectrum with multiple shades of identifiability.



This is a primer on how to distinguish different categories of data.

DEGREES OF IDENTIFIABILITY

Information containing direct and indirect identifiers.

PSEUDONYMOUS DATA

Information from which direct identifiers have been eliminated or transformed, but indirect identifiers remain intact.

DE-IDENTIFIED DATA

Direct and known indirect identifiers have been removed or manipulated to break the linkage to real world identities.

ANONYMOUS DATA

Direct and indirect identifiers have been removed or manipulated together with mathematical and technical guarantees to prevent re-identification.

	EXPLICITLY PERSONAL	POTENTIALLY IDENTIFIABLE	NOT READILY IDENTIFIABLE	KEY CODED	PSEUDONYMOUS	PROTECTED PSEUDONYMOUS	DE-IDENTIFIED	PROTECTED DE-IDENTIFIED	ANONYMOUS	AGGREGATED ANONYMOUS
DIRECT IDENTIFIERS Data that identifies a person without additional information or by linking to information in the public domain (e.g., name, SSN)	INTACT	PARTIALLY MASKED	PARTIALLY MASKED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED
INDIRECT IDENTIFIERS Data that identifies an individual indirectly. Helps connect pieces of information until an individual can be singled out (e.g., DOB, gender)	INTACT	INTACT	INTACT	INTACT	INTACT	INTACT	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED	ELIMINATED or TRANSFORMED
SAFEGUARDS and CONTROLS Technical, organizational and legal controls preventing employees, researchers or other third parties from re-identifying individuals	NOT RELEVANT due to nature of data	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	CONTROLS IN PLACE	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	LIMITED or NONE IN PLACE	CONTROLS IN PLACE	NOT RELEVANT due to nature of data	NOT RELEVANT due to high degree of data aggregation
SELECTED EXAMPLES	Name, address, phone number, SSN, government-issued ID (e.g., Jane Smith, 123 Main Street, 555-555-5555)	Unique device ID, license plate, medical record number, cookie, IP address (e.g., MAC address 68:A8:6D:35:65:03)	Same as Potentially Identifiable except data are also protected by safeguards and controls (e.g., hashed MAC addresses & legal representations)	Clinical or research datasets where only curator retains key (e.g., Jane Smith, diabetes, HgB 15.1 g/dl = Csrk123)	Unique, artificial pseudonyms replace direct identifiers (e.g., HIPAA Limited Datasets, John Doe = SL7T LX619Z) (unique sequence not used anywhere else)	Same as Pseudonymous, except data are also protected by safeguards and controls	Data are suppressed, generalized, perturbed, swapped, etc. (e.g., GPA: 3.2 = 3.0-3.5, gender: female = gender: male)	Same as De-Identified, except data are also protected by safeguards and controls	For example, noise is calibrated to a data set to hide whether an individual is present or not (differential privacy)	Very highly aggregated data (e.g., statistical data, census data, or population data that 52.6% of Washington, DC residents are women)